CLASSICAL CRYPTOGRAPHY

Classical ciphers are divided into three important classes.

MONOALPHABETIC CIPHERS: so called because letters of the plaintext alphabet are mapped into unique letters.

POLYALPHABETIC CIPHERS: so called because letters of the plaintext alphabet are mapped into letters of the ciphertext space depending on their position on the text.

STREAM CIPHERS: so called because a key stream is generated and used to encrypt a plaintext.

MONOALPHABETIC CIPHERS

SUBSTITUTION CIPHERS:

The keyspace is the set of permutations on $\{0, 1, 2, \ldots, 25\}$. For a given key π,

$$E_\pi(x_1 x_2 \cdots x_n) = \pi(x_1)\pi(x_2)\cdots\pi(x_n),$$

and

$$D_\pi(y_1 y_2 \cdots y_n) = \pi^{-1}(y_1)\pi^{-1}(y_2)\cdots\pi^{-1}(y_n),$$

Example 1: SHIFT (or CEASAR) CIPHERS:

This is a substitution cipher with permutation

$$x \to \pi(x) = x + b \bmod 26,$$

for some $0 \leq b \leq 25$.

Example 2: AFFINE CIPHERS:

This is a substitution cipher with permutation

$$x \to \pi(x) = ax + b \bmod 26,$$

for some $0 \leq a, b \leq 25$, and $\gcd(a, 26) = 1$.

EXAMPLE: CEASAR CIPHER

Given the encrypted message

$$L \;\; FDPH \;\; L \;\; VDZ \;\; L \;\; FRQTXHUHG$$

We have 26 possible keys. Therefore we can easily use exhaustive search and try all possible keys.

The permutation is

A	B	C	D	E	F	G	H	$\ldots$	Y	Z
d	e	f	g	h	i	j	k	$\ldots$	b	c

and the original plaintext message is

$$i \;\; came \;\; i \;\; saw \;\; i \;\; conquered$$

The CEASAR) CIPHER succumbs easily to **Ciphertext-Only** attacks.

METHODS OF CRYPTANALYSIS

Ciphertext-Only: The opponent possesses a string of ciphertext y.

Known Plaintext: The opponent possesses a string of plaintext x, and the corresponding ciphertext string y.

Chosen Plaintext: The opponent has obtained temporary access to the encryption machinery. He can choose a plaintext string x and construct the corresponding ciphertext string y.

Chosen Ciphertext: The opponent has obtained temporary access to the decryption machinery. He can choose a ciphertext string y and construct the corresponding plaintext string x.

Types of Security

There are two fundamentally different ways ciphers may be secure.

Unconditional Security: No matter how much computer power is available, the cipher cannot be broken

Computational Security: May mean one of two things.

1. Given limited computing resources (e.g., time needed for calculations is greater than age of universe), the cipher cannot be broken.

2. Provide evidence of computational security by reducing the security of the cryptosystem to some well-studied problem thought to be difficult (e.g., factoring). Such systems are called **provably secure**.

FREQUENCY ANALYSIS

In most languages letters occur in texts with different frequencies.

E.g., in English **E** is by far the most common letter.

We have tables of single, double, and triple letter frequencies derived from novels, newspapers, etc.

Frequencies are different for different languages

Single	Frequency	Double	Triple
E	.127	TH	THE
T	.091	HE	ING
A	.082	IN	AND
O	.075	ER	HER
I	.070	AN	ERE
N	.067	RE	ENT
S	.063	ED	THA
H	.061	ON	NTH

Cryptanalysis of Affine Ciphers

We consider **Ciphertext-Only** attacks.

Consider the ciphertext

$$FMNVEDKAPHFERBNDKRX$$
$$RSREFMORUDSDKDVSHVU$$
$$FEDKAPRKDLYEVLRHHRH$$

By tabulating we get the following frequencies in descending order

Letter	# of Occurrences
R	8
D	6
E	5
H	5
K	5
V	4
F	4

Making a guess

In each guess we choose two potential candidate letters

The most frequently occurring letters (in decreasing order of occurrence) are R, D, E, H, K.

Based on the frequency of occurrence of the letters we will make a guess and then "try to show that our guess makes sense".

Using our guess we will compute values a, b and confirm that $E_k(x) = ax + b \bmod 26$ is correct, for all letters x.

This involves solving a linear system consisting of two congruences with two unknowns (the unknowns are a, b).

Cryptanalysis of Affine Ciphers

1st guess: $R \to e$ and $D \to t$, i.e., $E_k(19) = 5, E_k(4) = 17$, where $E_k(x) = ax + b$ mod 26. It follows that $4a + b = 17$ mod $26, 19a + b = 5$ mod 26. Solving the system we get $a = 6, b = 19$, which is illegal since $\gcd(6, 26) > 1$.

2nd guess: $R \to e$ and $E \to t$. Proceeding as before this gives $a = 13$ which is again illegal because $\gcd(13, 26) > 1$.

3rd guess: $R \to e$ and $H \to t$. Proceeding as before we obtain $a = 3, b = 5$. The encryption function is $E_k(x) = 3x + 5$ mod 26. The decryption operation is easily seen to be $D_k(y) = 9y - 19$ mod 26. If we perform the decryption operation $D_k(y) = 9y - 19$ mod 26 on the ciphertext we obtain the plaintext: algorithms are quite general definitions of arithmetic processes.

Cryptanalysis of Substitution Ciphers

We consider **Ciphertext-Only** attacks.

We tabulate the frequency of the 26 letters in the ciphertext. We do the same for digrams and trigrams.

We guess that the most frequently occuring letter is e.

We use trial-and-error exhaustive analysis.

We look next at digrams and trigrams and make a guess which is consistent with our previous choices.

In each step we verify if our guess is "consistent" and whether or not it leads to meaningful plaintext.

Although there are 26! possible permutations of the 26 letters, this frequency analysis leads easily to conclusion even in plaintexts of length as little as 200 characters.

The Craft of Cryptanalysis

We will show how to decrypt the text

letter	frequency	letter	frequency
A	0	N	9
B	1	O	0
C	15	P	1
D	13	Q	4
E	7	R	10
F	11	S	3
G	1	T	2
H	4	U	5
I	5	V	5
J	11	W	8
K	1	X	6
L	0	Y	10
M	16	Z	20

```
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

to
our friend from Paris examined his empty glass with surprise, as if evaporation had taken place while he wasn't looking i poured some more wine and he settled back in his chair face tilted up towards the sun

Based on the frequency table we conjecture that $Z \to e$.

Other characters occuring more than ten times are C, D, F, J, M, R, Y.

We expect that they may decrypt to (a subset) of t, a, o, i, n, s, h, r.

Next we look at digrams of the form $-Z$ and $Z-$. By tabulation we find that DZ and ZW occur four times each. NZ and ZU three times each.

By looking at digram tables and since ZW occurs four times, but WZ not at all, and W occurs less than any other characters we guess $W \to d$.

Since DZ occurs four times and ZD occurs twice we could guess that $D \to r, s, t$ but is not clear which one.

Assuming $Z \rightarrow e$ $W \rightarrow d$ we look back at the ciphertext and notice that ZRW and RZW occur near the beginning and RW later on in the ciphertext.

Since RW occurs frequently and nd is a common digram we guess $R \rightarrow n$.

At this point we have completed the top picture.

Since NZ is a common digram while ZN is not another guess we make is $N \rightarrow h$

Assuming this is correct then segment $ne{-}ndhe$ suggests that $C \rightarrow a$

At this point we have completed the bottom picture.

```
------end---------e----ned---e------------
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

--------e----e---------n--d---en----e----e
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

-e---n------n------ed---e---e--ne-nd-e-e--
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

-ed-----n-----------e----ed-------d---e--n
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

```
------end-----a---e-a--nedh--e------a-----
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

h-------ea---e-a---a---nhad-a-en--a-e-h--e
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

he-a-n------n------ed---e---e--neandhe-e--
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

-ed-a---nh---ha---a-e----ed-----a-d--he--n
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR
```

Now consider M, the second most common cipherletter. The correspondence $KNM \to nh-$ suggests that $h-$ begins a word. So, probably, M is a vowel, i.e., $M \to i, o$.

Since ai is much more likely than ao the ciphertext digram CM suggests $M \to i$.

At this point we have completed the top picture.

Which letter is encrypted to o? Since o is a common letter we guess it must be one of D, F, J, Y. To avoid large strings of vowels it appears that $Y \to o$ is the best choice.

Of the remaining letters we conjecture $D, F, J \to r, s, t$. We guess $F \to r$ and $J \to t$.

At this point we have completed the bottom picture.

-----iend-----a-i-e-a-inedhi-e------a---i-
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

h-----i-ea-i-e-a---a-i-nhad-a-en--a-e-hi-e
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

he-a-n-----in-i----ed---e---e-ineandhe-e--
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

-ed-a--inhi--hai--a-e-i--ed-----a-d--he--n
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

o-r-riend-ro--arise-a-inedhise--t---ass-it
YIFQFMZRWQFYVECFMDZPCVMRZWNMDZVEJBTXCDDUMJ

hs-r-riseasi-e-a-orationhadta-en--ace-hi-e
NDIFEFMDZCDMQZKCEYFCJMYRNCWJCSZREXCHZUNMXZ

he-asnt-oo-in-i-o-redso-e-ore-ineandhesett
NZUCDRJXYYSMRTMEYIFZWDYVZVYFZUMRZCRWNZDZJJ

-ed-ac-inhischair-aceti-ted--to-ardsthes-n
XZWGCHSMRNMDHNCMFQCHZJMXJZWIEJYUCFWDJNZDIR

BEAUFORT CIPHER

Invented by Sir Francis Beaufort is the precursor to the Vigenere cipher.

Arrange the English alphabet on a 27×27 square.

The key is derived from a name, place, poem, etc, and agreed by both users, and applied as follows:

Find the first letter of the message text in the side column.

From the letter, trace horizontally across the table until finding the first letter of the key.

At the top of the column find the ciphertext letter.

Reverse these steps for decryption.

	ABCDEFGHIJKLMNOPQRSTUVWXYZ
A	ABCDEFGHIJKLMNOPQRSTUVWXYZ
B	BCDEFGHIJKLMNOPQRSTUVWXYZA
C	CDEFGHIJKLMNOPQRSTUVWXYZAB
D	DEFGHIJKLMNOPQRSTUVWXYZABC
E	EFGHIJKLMNOPQRSTUVWXYZABCD
F	FGHIJKLMNOPQRSTUVWXYZABCDE
G	GHIJKLMNOPQRSTUVWXYZABCDEF
H	HIJKLMNOPQRSTUVWXYZABCDEFG
I	IJKLMNOPQRSTUVWXYZABCDEFGH
J	JKLMNOPQRSTUVWXYZABCDEFGHI
K	KLMNOPQRSTUVWXYZABCDEFGHIJ
L	LMNOPQRSTUVWXYZABCDEFGHIJK
M	MNOPQRSTUVWXYZABCDEFGHIJKL
N	NOPQRSTUVWXYZABCDEFGHIJKLM
O	OPQRSTUVWXYZABCDEFGHIJKLMN
P	PQRSTUVWXYZABCDEFGHIJKLMNO
Q	QRSTUVWXYZABCDEFGHIJKLMNOP
R	RSTUVWXYZABCDEFGHIJKLMNOPQ
S	STUVWXYZABCDEFGHIJKLMNOPQR
T	TUVWXYZABCDEFGHIJKLMNOPQRS
U	UVWXYZABCDEFGHIJKLMNOPQRST
V	VWXYZABCDEFGHIJKLMNOPQRSTU
W	WXYZABCDEFGHIJKLMNOPQRSTUV
X	XYZABCDEFGHIJKLMNOPQRSTUVW
Y	YZABCDEFGHIJKLMNOPQRSTUVWX
Z	ZABCDEFGHIJKLMNOPQRSTUVWXY

POLYALPHABETIC CIPHERS

HILL CIPHERS:

The keyspace consists of the $n \times n$ invertible matrices. If k is an $n \times n$ invertible matrix then

$$E_k(x) = xk, \quad D_k(y) = yk^{-1}.$$

PERMUTATION CIPHERS:

For a given permutation π let $k_{i,j}^{\pi} = 1$ if $\pi(i) = j$, and $k_{i,j}^{\pi} = 0$, otherwise. This is a Hill cipher with matrix $k^{\pi} = (k_{i,j}^{\pi})$.

VIGENERE CIPHERS:

Let $k = (k_1, k_2, \ldots, k_n)$ be an n-element vector. Then

$$E_k(x) = x + k, \quad D_k(y) = y - k.$$

Cryptanalysis of Vigenere Ciphers

We consider **Ciphertext-Only** attacks.

First we consider two techniques to compute the block size m.

Kasiski's Observation: Two identical pieces of plaintext will be encrypted to the same ciphertext whenever their occurrence of plaintext is x positions apart, where $x = 0 \mod m$.

Kasiski's Test: Search the ciphertext for pairs of identical strings of length at least 3.

Record distances between the starting positions of the two segments.

If we record several such distances $d_1, d_2, \ldots$ then we can conjecture that $m | \gcd(d_1, d_2, \ldots)$.

Friedman's Test uses **Index of Coincidence:** Let $I_c(x)$ = probability that two random elements of the n-letter string x are identical.

Let $f_0, f_1, \ldots, f_{25}$ be the # of occurrence of $A, B, \ldots, Z$, respectively in the ciphertext.

$$I_c(x) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)} \qquad (1)$$

Now recall the frequency table

Single	Frequency
E	.127
T	.091
A	.082
O	.075
I	.070
N	.067
S	.063
H	.061

From this table we can get the expected value of $I_c(x)$; p_i^2 is the probability that in the string x both elements chosen at random are equal to the i-th letter.

Here, p_i be the expected probability of occurrence of the i-th letter in the English language.

Now English text and random strings differ:

Ciphertext	English Text	Random String
$I_c(x)$ in Formula (1)	$\sum_{i=0}^{25} p_i^2 =$ 0.065	$26(1/26)^2 =$ 0.038

Friedman's Method: Guess m. Partition the ciphertext $\mathbf{y}$ into m columns $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m$ each of length n/m.

If our guess m is the correct block size then $I_c(\mathbf{y}_i) \approx 0.065$, while if m is not correct then $I_c(\mathbf{y}_i) \approx 0.038$.

Example

Consider the ciphertext obtained from Vigenere cipher.

```
CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBW
RVXUOAKXAOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAK
LXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSRELXNJELX
VRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHR
ZBWELEKMSJIKNBHWRJGNMGJSGLXFEYPHAGNRBIEQJT
AMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQEBBI
PEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHP
WQAIIWXNRMGWOIIFKEE
```

We will analyze this ciphertext. Our method will be the following.

1. We will make a guess for the block size and subsequently use, (a) Kasiski's method, and (b) Friedman's method to compute m. Both methods will give the same value for the block size.

2. Assuming now that m is known we will use frequency analysis and the mutual index of coincidence to decipher the text.

Kasiski's Method: The cipher text string CHR occurs in four places in ciphertext and at the positions:

$$1, 166, 236, 286.$$

The distances from the first occurrence are $165, 235, 285$.

Since $\gcd(165, 235, 285) = 5$ our guess for the block size is 5.

Friedman's Method: We compute the index of coincidence for the given ciphertext.

For each guess m, we have blocks of size n/m. So we have block size 5 as the best guess.

m	I_c
1	0.045
2	0.046, 0.041
3	0.043, 0.050, 0.047
4	0.042, 0.039, 0.046, 0.040
5	0.063, 0.068, 0.069, 0.061, 0.072

Assume we know the block size m

Mutual index of coincidence: $MI_c(x, x')$ is the probability that a random element of x is identical to a random element of x'.

Let $f_0, f_1, \ldots, f_{25}, f'_0, f'_1, \ldots, f'_{25}$ be the frequency of occurrence of $A, B, \ldots, Z$, respectively in the strings x, x'. So we have the formula.

$$MI_c(x, x') = \frac{\sum_{i=0}^{25} f_i f'_i}{nn'} \qquad (2)$$

Since we know m we can partition the ciphertext y into m blocks $y_1, \ldots, y_m$.

Assuming $K = (k_1, k_2, \ldots, k_m)$ is the key being used we can estimate $MI_c(\mathbf{y}_i, \mathbf{y}_j)$.

Take a random character in $\mathbf{y}_i$ and a random character in $\mathbf{y}_j$. The probability that both characters are A is $p_{-k_i} p_{-k_j}$. The probability that both characters are B is $p_{1-k_i} p_{1-k_j}$, etc. Hence,

$$MI_c(\mathbf{y}_i, \mathbf{y}_j) \approx \sum_{h=0}^{25} p_{h-k_i} p_{h-k_j} = \sum_{h=0}^{25} p_h p_{h+k_i-k_j}.$$

Note that the mutual coincidence depends on the relative shift $k_i - k_j$. Moreover, the mutual coincidence of the relative shift $k_i - k_j$ is the same with the mutual coincidence of the relative shift $k_j - k_i$.

Relative Shift	Expected value of MI_c
0	0.065
1	0.039
2	0.032
3	0.034
4	0.044
5	0.033
6	0.036
7	0.039
8	0.034
9	0.034
10	0.038
11	0.045
12	0.039
13	0.043

Next we tabulate the relative shifts between 0 and 13.

The mutual coincidences vary from 0.031 to 0.045 if the relative shift is nonzero

The mutual coincidence is 0.065 if the relative shift is 0.

This observation is being used to formulate a likely guess for the shift $\ell = k_i - k_j$.

Now fix $\mathbf{y}_i$. Let $y_j^0, y_j^1, y_j^2, \ldots$ be the result of encrypting y_j by $e_0, e_1, e_2, \ldots$.

Tabulate the mutual indices $MI_c(\mathbf{y}_i, \mathbf{y}_i^g)$.

When $g = \ell$ the mutual index of coincidence should be 0.065.

When $g \neq \ell$ the mutual index of coincidence should be between 0.031 and 0.045.

Tabulate, by computer, the 260 values

$$MI_c(y_i, y_j^g),\ 1 \le i < j \le 5, 0 \le g \le 25$$

For each (i, j) look at the values of $MI_c(y_i, y_j^g)$ which are close to 0.065. If the value is unique we conjecture it is the value of the relative shift.

Six such potential values are boxed in the table.

Pair	Relative Shift
$(1,2)$	9
$(1,5)$	16
$(2,3)$	13
$(2,5)$	7
$(3,5)$	20
$(4,5)$	11

The resulting equations are

$$\begin{array}{ll} k_1 - k_2 = 9 & k_1 - k_5 = 16 \\ k_2 - k_3 = 13 & k_2 - k_5 = 7 \\ k_3 - k_5 = 20 & k_4 - k_5 = 11 \end{array}$$

The key is $(k_1, k_1+17, k_1+4, k_1+21, k_1+10)$, for some $k_1 \le 25$, which is a cyclic shift of $AREVK$. The key is easily found: $JANET$.

Observed Mutual Indices of Coincidence

i	j	value of $MI_c(\mathbf{y}_i, \mathbf{y}_j^g)$								
1	2	.028	.027	.028	.034	.039	.037	.026	.025	.052
		.068	.044	.026	.037	.043	.037	.043	.037	.028
		.041	.041	.034	.037	.051	.045	.042	.036	
1	3	.039	.033	.040	.034	.028	.053	.048	.033	.029
		.056	.050	.045	.039	.040	.036	.037	.032	.027
		.037	.036	.031	.037	.055	.029	.024	.037	
1	4	.034	.043	.025	.027	.038	.049	.040	.032	.029
		.034	.039	.044	.044	.034	.039	.045	.044	.037
		.055	.047	.032	.027	.039	.037	.039	.035	
1	5	.043	.033	.028	.046	.043	.044	.039	.031	.026
		.030	.036	.040	.041	.024	.019	.048	.070	.044
		.028	.038	.044	.043	.047	.033	.026	.046	
2	3	.046	.048	.041	.032	.036	.035	.036	.030	.024
		.039	.034	.029	.040	.067	.041	.033	.037	.045
		.033	.033	.027	.033	.045	.052	.042	.030	
2	4	.046	.034	.043	.044	.034	.031	.040	.045	.040
		.048	.044	.033	.024	.028	.042	.039	.026	.034
		.050	.035	.032	.040	.056	.043	.028	.028	
2	5	.033	.033	.036	.046	.026	.018	.043	.080	.050
		.029	.031	.045	.039	.037	.027	.026	.031	.039
		.040	.037	.041	.046	.045	.043	.035	.030	
3	4	.038	.036	.040	.033	.036	.060	.035	.041	.029
		.058	.035	.035	.034	.053	.030	.032	.035	.036
		.036	.028	.046	.032	.051	.032	.034	.030	
3	5	.035	.034	.034	.036	.030	.043	.043	.050	.025
		.041	.051	.050	.035	.032	.033	.033	.052	.031
		.027	.030	.072	.035	.034	.032	.043	.027	
4	5	.052	.038	.033	.038	.041	.043	.037	.048	.028
		.028	.036	.061	.033	.033	.032	.052	.034	.027
		.039	.043	.033	.027	.030	.039	.048	.035	

The almond tree was in tentative blossom. The days were longer often ending with magnificent evenings of corrugated pink skies. The hunting season was over ...

Cryptanalysis of Hill Ciphers

We consider **Known Plaintext** attacks. First, assume we know the "block" size n.

Assume we have n plaintext-ciphertext pairs

$$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n).$$

If k is the unknown $n \times n$ matrix then

$$Y_i = X_i k, \text{ for all } i = 1, 2, \ldots, n.$$

If $X = [X_1, X_2, \ldots, X_n]$, and $Y = [Y_1, Y_2, \ldots, Y_n]$ are the corresponding row matrices then $Y = Xk$ and hence $k = X^{-1}Y$.

If X is not invertible then we must look for another set of plaintext-ciphertexts.

If n is not known then we could simply try by exhaustive search $n = 2, 3, \ldots$ until the correct value is found.

STREAM CIPHERS

In addition to the plaintext, ciphertext and key spaces, stream ciphers are endowed with a key stream alphabet L and a key stream generator $F = \{f_1, f_2, \ldots\}$, where

$$f_i : K \times P^{i-1} \rightarrow L$$

A key stream $\ell_1, \ell_2, \ldots$ is generated and used to encrypt a plaintext $x = x_1 x_2 \cdots$ according to the rule:

$$E_{\ell_1}(x_1) E_{\ell_2}(x_2) \cdots,$$

where

$$\ell_i = f_i(k, x_1, \ldots, x_{i-1}),$$

and $k \in K$.

For each $\ell \in L$ the encryption and decryption functions E_ℓ, D_ℓ satisfy $D_\ell(E_\ell(p) = p$, for all $p \in P$.

We can think of block ciphers as special cases of stream ciphers where the key stream is constant.

A stream cipher is **synchronous** if the keystream is independent of the plaintext, i.e. keystream is generated only as a function of the key (k is called the "seed").

A stream cipher is **periodic** with **period** d if $\ell_{i+d} = \ell_i$, for all $i \geq 1$.

The Vigenere cipher with keyword length m is a periodic stream cipher with keyword length m.

Stream ciphers often described in binary $0, 1$ alphabets: e.g., $E_\ell(x) = x + \ell \bmod 2, D_\ell(x) = x + \ell \bmod 2$.

We can generate a **synchronous** keystream via a linear recurrence relation of degree m:

$$\ell_{i+m} = \sum_{j=0}^{m-1} c_j \ell_{i+j} \text{ mod } 2,$$

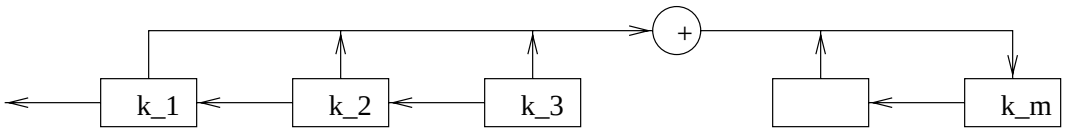
where we start with $k_1, k_2, \ldots k_m$, set $\ell_i = k_i$, for $i \leq m$, and $c_0, c_1, \ldots, c_{m-1}$ are predetermined constants.

Example: Assume $m = 4$ and the keystream is generated via $\ell_{i+4} = \ell_i + \ell_{i+1}$ mod 2, where $i \geq 1$.

Keystream initialized with any nonzero string, e.g. starting with $(1, 0, 0, 0)$ we obtain the stream:

$$1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, \ldots$$

An appealing aspect of keystream generation is that it can be produced efficiently in hardware using a Linear Feedback Shift Register (LFSR).



Following operations performed at each time unit:

k_1 is tapped as the next keystream bit

$k_2, k_3, \ldots, k_m$ would each be shifted one stage to the left.

the new value k_m would be computed to be

$$\sum_{j=0}^{m-1} c_j k_{j+1},$$

where the c_j's are constants 0 or 1 specified by the register.

Example of LFSR

Consider the key $k = 8$ and the plaintext:

rendezvouz

Plaintext is converted to integers.

The keystream starts with key 8, and generated with the **autokey cipher** which shifts the plaintext characters by one position and ciphertext is obtained by adding the columns modulo 26:

$\ell_1 = k, \ell_i = x_{i-1}$
$E_\ell(x) = x + \ell \bmod 26, D_\ell(y) = y - \ell \bmod 26$

We have the following encryption

plain:	r	e	n	d	e	z	v	o	u	s
keys:	17	4	13	3	4	25	21	14	20	18
	8	17	4	13	3	4	25	21	14	20
cipher:	25	21	17	16	7	3	20	22	8	12
	Z	V	R	Q	H	D	U	J	I	M

For decryption we reverse the previous steps.

Cryptanalysis of LFSR: We consider **Known Plaintext**) attacks.

Ciphertext is the sum modulo 2 of plaintext and the keystream:

$$\ell_{i+m} = \sum_{j=0}^{m-1} c_j \ell_{i+j} \bmod 2,$$

which is a linear equation in m unknowns. In matrix form it can be written as $(\ell_{m+1}, \ldots, \ell_{2m}) =$

$$(c_0, \ldots, c_{m-1}) \begin{pmatrix} \ell_1 & \ell_2 & \ldots & \ell_m \\ \ell_2 & \ell_3 & \ldots & \ell_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ \ell_m & \ell_{m+1} & \ldots & \ell_{2m-1} \end{pmatrix}$$

which implies that $(c_0, \ldots, c_{m-1}) =$

$$(\ell_{m+1}, \ldots, \ell_{2m}) \begin{pmatrix} \ell_1 & \ell_2 & \ldots & \ell_m \\ \ell_2 & \ell_3 & \ldots & \ell_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ \ell_m & \ell_{m+1} & \ldots & \ell_{2m-1} \end{pmatrix}^{-1}$$

Thus for a given block size m if the stream is at least $2m$ bits long we can "break" the stream cipher easily, provided the matrix is invertible.